NAMRL 1282

DETERMINING VISUAL ACUITY THRESHOLDS:

A Simulation Study of Stimulus Presentation Strategies

David J. Blower

August 1981

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY
PENSACOLA, FLORIDA

Approved for public release; distribution unlimited.

DETERMINING VISUAL ACUITY THRESHOLDS:

A Simulation Study of Stimulus Presentation Strategies

David J. Blower

Naval Medical Research and Development Command
ZF51.524-004-2011

Naval Air Systems Command
W 43-13.8881

Reviewed by:

Ashton Graybiel, M.D.
Chief Scientific Advisor

Approved and Released by:

Captain W. M. Houk, MC, USN
Commanding Officer

10 August 1981

Naval Aerospace Medical Research Laboratory
Naval Air Station
Pensacola, Florida 32508

SUMMARY PAGE

THE PROBLEM

There are a number of tests in the Vision Test Battery where we want to accurately and economically determine the threshold stimulus. This is a two pronged problem which first depends upon how one chooses to present the visual stimuli and, secondly, upon how one chooses to analyze the data after the stimuli have been presented. This paper addresses the first of these problems through a computer simulation of two strategies for presenting visual stimuli. Three basic questions were posed in this study: 1) How close can the classical up-down method of presenting stimuli come to the "true" threshold in a four alternative forced choice task? 2) How variable is this estimate of the theshold as a function of trial number? 3) Can another method of stimulus presentation be devised which will provide better results than the up-down method according to the criteria laid down in 1) and 2) above?

FINDINGS

A true acuity threshold was defined as the mean of a normal distribution, and the cumulative normal distribution was assumed to be representative of a psychometric function for one particular acuity test in the Vision Test Battery (VTB). One hundred simulation runs using the up-down method of presenting stimuli showed that, on the average, this method underestimated the true acuity threshold. In addition, the variability of the threshold estimator was determined for the up-down method as a function of trial number. One hundred simulation runs of a new method of presenting stimuli were also conducted. This method provided an estimate which, on the average, was closer to the true threshold and which was less variable than the estimate provided by the up and down method. This new method also proved to be superior when the slope parameter of the psychometric function was varied over a large range.

ACKNOWLEDGMENTS

I should like to acknowledge Captain James E. Goodson, MSC, USN, (Ph.D.) for his help in identifying the problem addressed in this paper and for posing questions which, to some extent, guided my inquiry. I should also like to extend my thanks to the rest of the members of the Vision Sciences Division for their encouragement, and for their discussion of the ideas contained in this paper.

INTRODUCTION

One of the goals of visual psychophysics is the determination of the threshold stimulus. This is usually defined as that stimulus which the subject can detect fifty percent of the time. The Vision Sciences Division of NAMRL has developed a Vision Test Battery (VTB) comprised of a large number of tests designed to measure various aspects of visual performance. Visual performance in the VTB is ultimately intended to be correlated with success in tactical air combat maneuvers, such as are conducted at the Air Combat Maneuvering Range.

There are a number of tests in the VTB where we want to accurately and economically determine an acuity threshold. For example, one of the tests in the VTB is the far, central, high contrast acuity test.

In this test the subject attempts to correctly locate a gap in a Landolt C ring. There are four possible locations of the gap in the Landolt ring: up, down, right, and left. The subject indicates his choice of where the gap is located by moving a joystick in a direction corresponding to the location of the gap. The Landolt C ring is located 5.5 meters (18 feet) from the subject in his central visual field. The contrast of the Landolt C is 100% against a background lighting brightness of 343 cd/m^2.

We have a set of twenty precisely measured Landolt C rings whose gaps range from 1.51 minutes of visual angle (mva) to 0.18 mva. These are the stimuli which we will present to the subject in order to determine his acuity threshold in this far, central, high contrast test. The threshold stimulus as measured in minutes of visual angle will serve as a score for a subject in this test, and, as stated before, we would like to obtain it as accurately as we can in the shortest time possible because of the many other tests still awaiting the subject.

PLANNED PRESENTATION vs. SEQUENTIAL PRESENTATION OF THE STIMULI

There are two main strategies to presenting the stimuli. One strategy is to use a planned presentation of stimuli such as the method of constant stimuli. For example, five or six Landolt C rings of differing gap sizes could be chosen as the stimuli to present in advance of any response by the subject. These stimuli could be shown to the subject twenty times each in a random order and an estimate made of his acuity threshold.

A different strategy which utilizes sequential presentation of stimuli could be employed. This approach uses the subject's responses to guide the selection of stimuli for future presentation. An advantage of the sequential presentation strategy is that it tends to concentrate the presentation of stimuli in the region of most interest on the psychometric curve. It has been shown (Cornsweet (3)) that such sequential methods can lead to estimates of threshold acuity which are as precise as those obtained with a larger number of trials using planned presentation strategy.

In the planned presentation strategy, such as the method of constant stimuli, our initial guess as to the location of the threshold may be in error, resulting in a large number of observations which give little or no information as to the location of the threshold. In this sense, it is inefficient when compared with sequential presentation of stimuli where we can choose the next stimulus to be located close to the threshold.

One method of presenting stimuli in a sequential fashion is the up and down method first proposed by Dixon and Mood (4). This sequential design operates in a very easy to understand manner. If a subject responds correctly to a stimulus of a given gap size on trial n, then on trial n+1 we present him with the next smaller stimulus. If the subject responds incorrectly to a stimulus of a given gap size on trial n, then on trial n+1 we present him with the next larger stimulus. As we proceed in this fashion, the stimuli presented are those which are close to the threshold stimulus. However, Blower (1) has shown that if the up and down method is used with a four alternative forced choice method of responding, the stimuli presented will have a tendency to be below the threshold stimulus. As a consequence, an estimate of the threshold stimulus calculated from such data may be an underestimation of the true threshold.

There are other sequential presentation strategies besides the two discussed in this paper. There is the PEST technique of Taylor and Creelman (8), the transformed up and down methods (Levitt (5)), a maximum likelihood technique (Pentland (7)), and other variations on these basic sequential themes.

MAIN OBJECTIVES

The purpose of this paper is to report on a simulation study of certain statistical features of the up and down method. There were three basic questions we wanted to answer before proceeding with a sequential experimental design for presenting stimuli to determine acuity thresholds. They were:

1) How close can the up and down method come to estimating a known threshold when a four alternative forced choice task is employed?

2) What is the error in the estimate of the threshold as a function of the number of stimulus presentations?

3) Is there a better method of presenting stimuli in a sequential manner using the criteria of 1) and 2) above?

In order to answer these questions a simulation study was carried out. A description of the assumptions underlying the simulation is contained in the next section.

THE SIMULATION MODEL

A cumulative normal distribution was assumed to be a good model for the psychometric function relating probability of correct detection to the size of the gap in the Landolt C ring. The parameters of the normal distribution function for the far, central, high contrast test were taken to be μ (population mean) = 0.50 mva, and σ (population standard deviation) = 0.08 mva. In other words, the gap size giving rise to 50% probability of correct detection, and, therefore, the true threshold, is half a minute of visual angle. Since σ = 0.08 mva, the model stipulates that, for example, at a gap size of 0.66 mva there should be about a 98% chance of correct detection. These parameter values were chosen to correspond to reasonable estimates of the location and slope of a psychometric curve based on previous data (Lythgoe (6)) for the far, central, high contrast test.

Figure 1 shows the theoretical psychometric curve. The probability of correct detection lies along the y-axis, and the size of the gap in the Landolt ring lies along the x-axis. Six of the twenty gap sizes available for the far, central, high contrast test are positioned on this psychometric curve. Gap sizes #1 through #8 are all large enough so that we would expect the probability of a correct detection to be nearly 100% under the terms of the model. Gap sizes #9, #10, #11, and #12 are all located on the steep portion of the psychometric curve with correspondingly

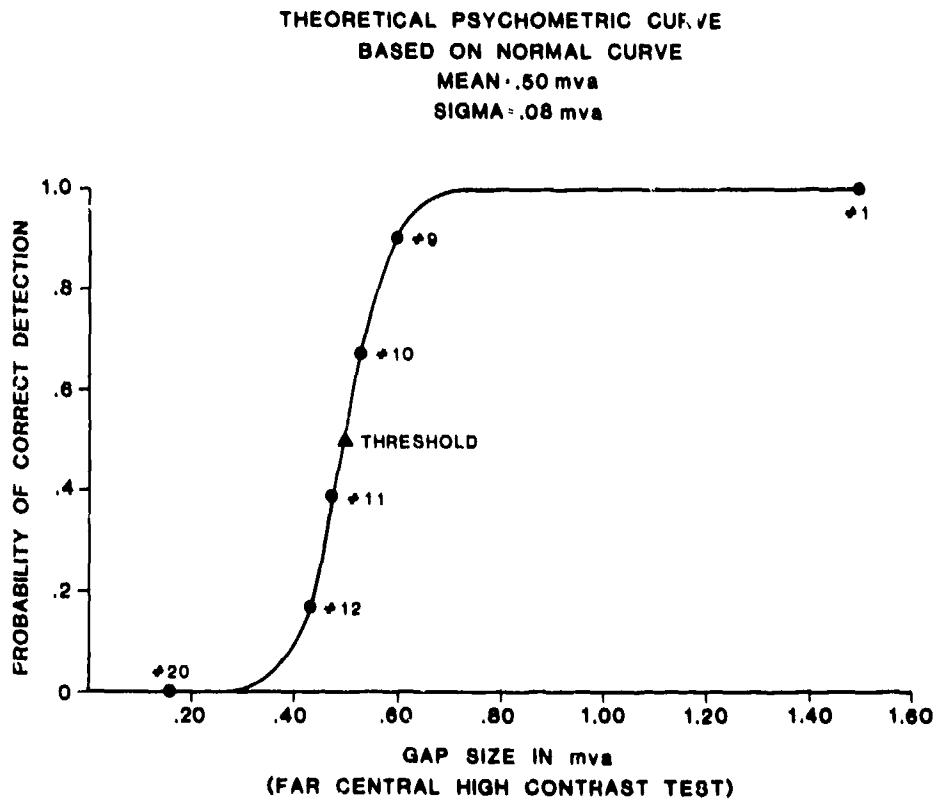

Figure 1

The theoretical psychometric curve used in the simulation model. This curve is a cumulative normal distribution function with μ = 0.5000 mva and σ = 0.08 mva. The threshold stimulus is a Landolt C ring with a gap size of 0.5000 mva.

decreasing probabilities of detection. The true threshold (0.50 mva) is seen to fall about midway between gap size #10 and #11. Gap sizes #13 through #20 are so small that we rarely expect them to be detected.

Now, since the subject is engaged in a four alternative forced choice task, the probability of a correct response will be greater than the probability of a correct detection for a given gap size. This correction for guessing is shown in Table I for one particular gap size. We see that while the probability of a correct detection for gap size #10 is 67%, the probability of a correct response is 75%.

Table I

Calculation of the probability of a correct response to a given gap size for the assumed psychometric curve.

A. Let a normal curve with a mean equal to 0.50 mva and a standard deviation equal to 0.08 mva serve as a model of the psychometric function.

B. Let's choose gap size #10 whose gap size is 0.534962 mva when projected on the far screen.

C. Calculation of z score for gap size #10.

$$z = \frac{0.534962 - 0.5}{0.08} = 0.4370$$

$$P(z \leq 0.4370) = 0.6689$$

D. Correction for guessing in a four alternative forced choice task.

$$P(\text{correct response}) = 0.25\ (1 - 0.6689) + 0.6689 = 0.7517$$

The probability of a correct response to all twenty gap sizes is given in Table II. We notice from this table that the small gap sizes still have a 25% probability of a correct response even though they cannot be detected at all. We also note that because of the four alternative forced choice nature of the task, the threshold stimulus is that gap size which leads to a 62.5% probability of correct response.

To this point we have explained the model of the assumed true underlying psychometric curve. Two additional assumptions were made in this simulation. First, we assumed that the parameters of the psychometric function did not change as a function of trial number. No shifts in either the value of the mean or standard deviation were allowed during the course of a simulation run. Secondly, we assumed that the probability of a correct response on a trial is independent of responses made on previous trials. We are assuming that the subject is responding solely to sensory information on each trial and not responding on the basis of psychological factors such as memory for previous responses, desire to seek or avoid certain response patterns, and so on.

Table II

The probability of a correct response to each of the twenty different gap sizes used in the far, central, high contrast, acuity test. The probability of a correct response is based on a psychometric curve with a mean of 0.50 mva and a standard deviation of 0.08 mva and a four alternative forced choice task.

Slide Number	Gap Size in mva	Prob. of correct response
1	1.513100	0.9999+
2	1.348019	0.9999+
3	1.200948	0.9999+
4	1.069923	0.9999+
5	0.953193	0.9999+
6	0.849199	0.9999+
7	0.756550	0.9995
8	0.674010	0.9889
9	0.600474	0.9215
10	0.534962	0.7517
11	0.476597	0.5388
12	0.424599	0.3798
13	0.378275	0.2982
14	0.337005	0.2658
15	0.300237	0.2547
16	0.26748[illegible]	0.2514
17	0.238298	0.2505
18	0.212300	0.2502
19	0.189138	0.2500+
20	0.168502	0.2500+

SIMULATION RESULTS OF THE UP AND DOWN METHOD

Figure 2 illustrates one simulation run of the up and down method using the above model. The y-axis indicates the twenty gap sizes. Remember that gap size #1 is the largest and gap size #20 is the smallest. The x-axis indicates the number of trials in the simulation run. There were 100 trials conducted for each run.

We start the simulation run by presenting gap size #1 on the first trial. An "x" indicates a correct response by the simulated subject, and an "0" indicates an incorrect response. This graph illustrates clearly that when the simulated subject makes a correct response, the next smaller gap size s presented, and when the simulated subject makes an incorrect response, the next larger gap size is presented.

Since the probability of a correct response to the larger gap sizes is quite high, we observe an initial run of correct responses. We then observe the peaks and valleys of the up and down process as it seeks to converge upon the mean of the normal distribution, or, equivalently. the acuity threshold. The true threshold of 0.50 mva, which falls between gap sizes #10 and #11 is indicated in the figure by the dark line.

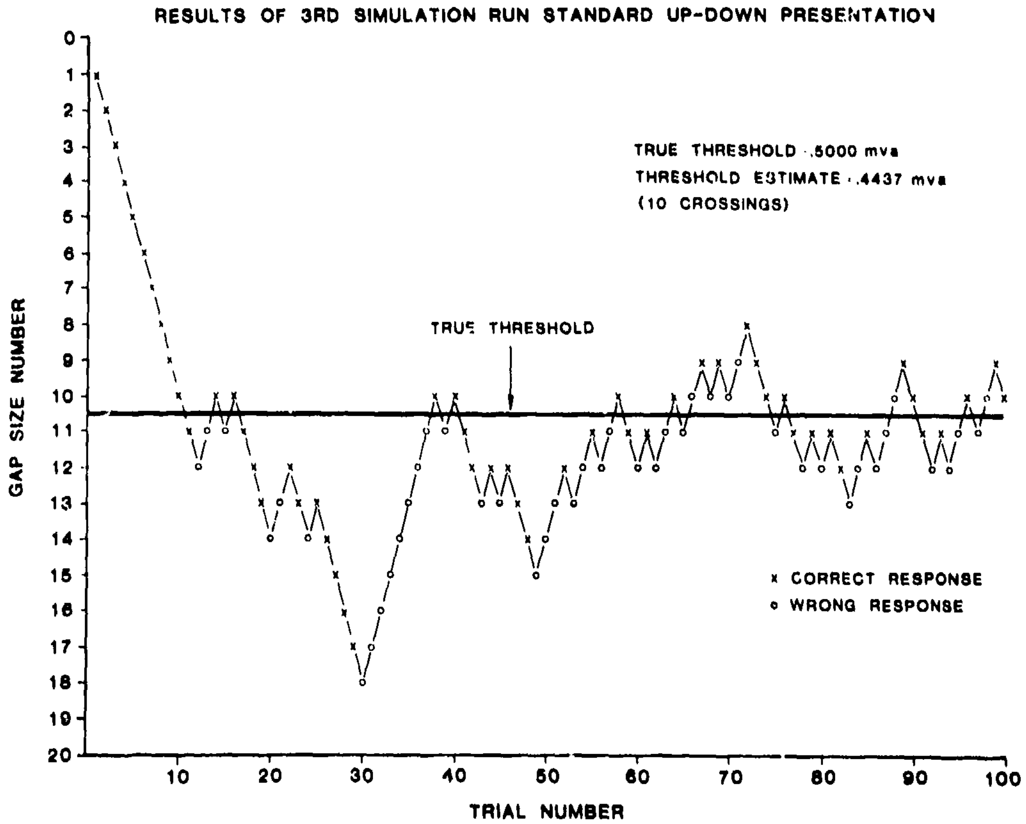

Figure 2

The gap size numbers presented in 100 trials of a simulation run of the up-down method. The probability of a correct response or wrong response on each trial is determined by the theoretical psychometric curve.

The nature of the stimulus presentation procedure by the up and down method has now been clearly explained. However, there still remains the task of forming an estimate of the threshold from the data. There are a number of ways to do this. Dixon and Mood (4), Brownlee et al. (2), and Wetherill et al. (9) reported different ways of estimating the threshold from data generated by the up and down method. Our analysis employs two estimates, one closely related to the Brownlee et al. average level estimator, and the second one similar to Wetherill's average of the peaks and valleys occurring within a run.

Our first method of estimating the threshold from the data generated by the simulated run of the up and down method is simply to average the gap sizes presented after the first ten trials. These first ten trials serve as practice trials in the actual experimental setting. This method of estimation will be called the "all data" method. The second method is to take the average of the two gap sizes where a wrong response to correct response reversal occurs. More precisely, we average the gap sizes of gap size #j and #j-1 if the following pattern occurs: trial n-2 = j, trial n-1 = j-1 and trial n = j. In the future this will simply be called a crossing. A selected number of these crossings will then be averaged to form an estimate of the threshold. For example, we see that the first crossing in Figure 2 (after the ten practice trials) occurs at trials 13 and 14. (Trial 12 = #11, trial 13 = #10, and trial 14 = #11). We would then average the gap sizes of #10 and #11 ((0.534962 + 0.476597) $\div$ 2 = 0.505780 mva) as one crossing. Figure 2 shows that if we average ten of these crossings our estimate of the true threshold of 0.50 mva is 0.4437 mva.

Figure 3 illustrates another realization of a simulation run of the up and down method using our model. We can observe the different patterns which arise from the stochastic nature of the up and down method by comparing Figures 2 and 3. The threshold estimate from ten crossings is 0.5022 mva for the data in Figure 3. This estimate is much closer to the true threshold than the estimate calculated from the data in Figure 2, and also gives some idea of the variability of estimates which can arise from the same underlying psychometric model.

One hundred simulation runs in all were conducted. Figures 2 and 3 are two representative examples from these one hundred runs. Figure 4 presents a plot of the average estimate of the true threshold over these one hundred runs for both the "all data" and "crossing" methods of estimation. The "all data estimator" can be plotted directly as a function of trial number. The "crossing estimator" is also plotted as a function of trial number, but the number of trials taken for a fixed number of crossings is itself a random variable. Therefore, the data points on the "crossing estimator" curve are taken to be the average number of trials for that number of crossings. For example, it takes, on the average, about forty trials to achieve ten crossings.

Figure 4 reveals that both methods underestimate the true threshold, with the "all data estimator" poorer than the "crossing estimator." Both methods seem to reach an asymptote at around forty trials (ten crossings) with a value of 0.48 mva for the "crossing estimator" and a value of about 0.46 mva for the "all data estimator." This underestimation of the true threshold arises because the presentation of stimuli in a four alternative forced choice task is shifted to stimuli with smaller gap sizes than the threshold gap size (Blower (1)).

Figure 4 answers one of the questions posed in the Introduction. It shows how close two estimators from the up and down method can come to a known threshold.

Perhaps more important than knowing how close these estimators can come to a true threshold is knowledge about the variability of the estimates. It is generally desirable to have an estimate with as small a variance as possible.

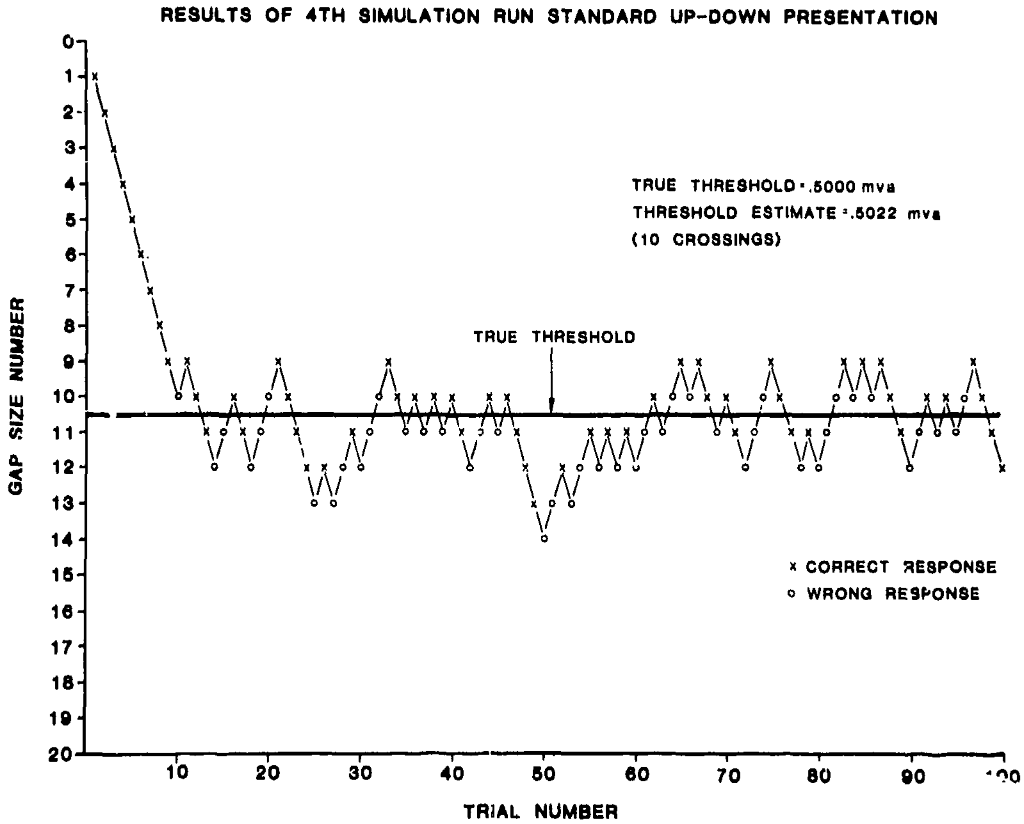

Figure 3

Another example of a 100 trial simulation run of the up-down method. Comparing Figure 2 with Figure 3 gives some idea of the different outcomes possible in the sequential presentation of stimuli even when the same psychometric function underlies both runs.

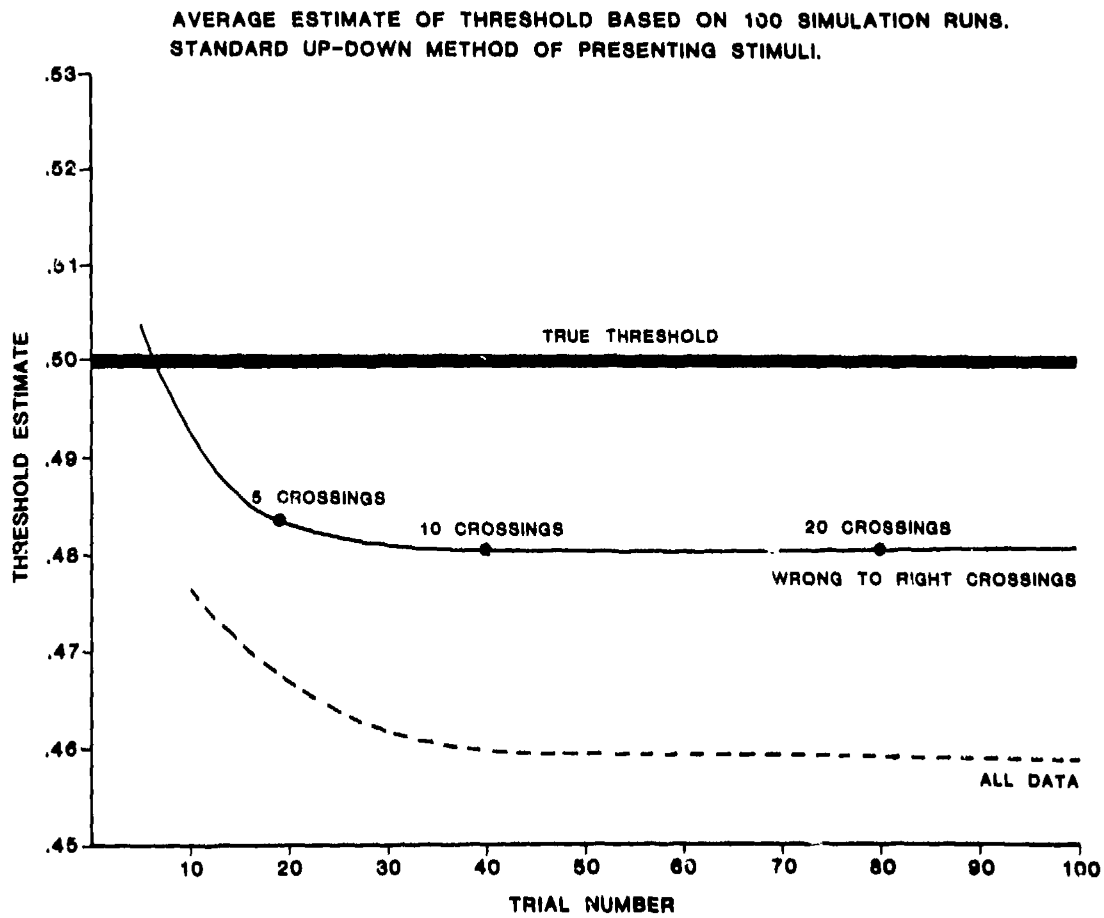

Figure 4

The two curves illustrate the extent to which two different estimation procedures underestimate the true threshold stimulus when the up-down method has been used with four alternative forced choice responding. Each point on the curves is the average of one hundred simulation runs.

Figure 5 sheds some light on this question. Figure 5 plots the average of the crossing estimator from the one hundred simulation runs as a function of the number of crossings. In addition it gives the range of estimates of threshold acuity to be expected for a given number of crossings. We assume that the estimator is normally distributed to find the upper and lower limits which contain 95% of the distribution. We do this by taking the average of the one hundred estimates plus and minus two times the standard deviation of the one hundred estimates.

As an example, the average of the threshold acuity estimates for one hundred simulation runs at ten crossings is 0.4805 mva. The standard deviation of these one hundred estimates is 0.0317 mva. We, therefore, expect about 95% of the distribution of this estimator to fall within the range of 0.4171 mva to 0.5439 mva.

We observe the entirely expected pattern of increased accuracy of our estimator as the number of crossings increases. However, the variability declines rather slowly with the number of crossings, and a rather unpalatable error still remains for as many as twenty crossings.

A NEW METHOD FOR PRESENTING STIMULI IN A SEQUENTIAL MANNER: THE BRACKET METHOD

One of the objectives of this research effort was to find an alternative method of presenting stimuli in a sequential manner which would enable us to estimate the true threshold acuity with greater accuracy and with lower variability than shown in Figure 5 for the up and down method.

A new method, called the "bracket method," will now be described which meets these objectives. The bracket method is most easily understood by referring to Table III. Table III is a combination of a numerical example and a computer flow-chart of how the bracket method operates.

The first ten practice trials are conducted according to the up and down method as previously discussed. Table III begins, therefore, with trial 11. The gap size presented on trial 11 is that gap size which was last presented by the up and down method on trial 10. Let's say that gap size #10 is presented on trial 11. The column labelled "N" lists the total number of times that this particular gap size was presented, while the column labelled "K" lists the total number of times that particular gap size was responded to correctly. The column labelled "delta" lists the absolute value of the difference between 62.5% and the total percentage correct for that gap size at that trial. 62.5% is the target point on the psychometric curve we are trying to locate. (50% correct detection, the threshold percentage, translates to 62.5% correct responses in a four alternative forced choice task.)

As long as delta is decreasing, we present the same gap size on the next trial. When delta is no longer decreasing, we check to see whether we were above the target probability of 62.5% or below it. If we were above 62.5%, we present the next smaller gap size on the succeeding trial; if below 62.5%, we present the next larger gap size on the succeeding trial.

Table III indicates how we would continue in this manner for a given number of trials. This method tries to converge to the presentation of two adjacent gap sizes which bracket the threshold gap size, and then to keep on presenting these bracketing gap sizes for the duration of the run.

In Table III we observe the successful operation of this method as it oscillates between gap sizes #10 and #11 from trial 17 on. Since the gap size of #10 is 0.53 mva and the gap size of #11 is 0.47 mva, we see that these two stimuli do indeed bracket the true threshold of 0.50 mva.

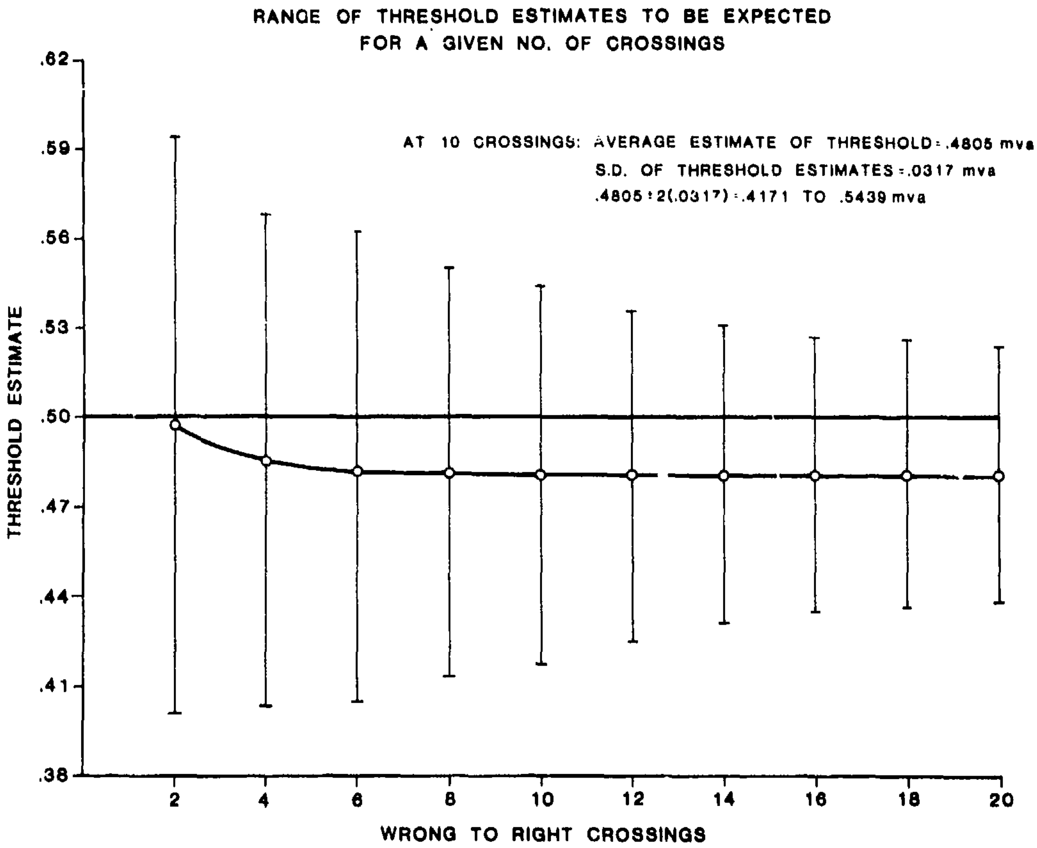

Figure 5

An indication of the error to be attached to a threshold estimate when using the "crossing estimator" for data generated by the up-down method. The error bars are plus and minus two standard deviations as calculated from the sample of threshold estimates from 100 simulation runs.

Table III

A numerical example of the algorithm used in the bracket method to present stimuli.

Trial #	Gap Size #	N	Correct Response?	K	% Correct	Greater than 62.5%?	Delta	Is Delta Decreasing?
11	10 (0.535 mva)	1	YES	1	100.0	YES	0.375	YES
12	10	2	NO	1	50.0	NO	0.125	YES
13	10	3	YES	2	66.7	YES	0.042	YES
14	10	4	NO	2	50.0	NO	0.125	NO

Since delta is no longer decreasing and we are below the desired percentage of 62.5, we present next larger target. Delta reset to 1.00.

15	9 (0.600 mva)	1	YES	1	100.0	YES	0.375	YES
16	9	2	YES	2	100.0	YES	0.375	NO

Since delta is no longer decreasing and we are above the desired percentage of 62.5, we present next smaller target. Delta reset to 1.00.

17	10	5	YES	3	60.0	NO	0.025	YES
18	10	6	YES	4	66.7	YES	0.042	NO

Since delta is no longer decreasing and we are above the desired percentage, we present next smaller target. Delta reset to 1.00.

19	11 (0.477 mva)	1	NO	0	0.0	NO	0.625	YES
20	11	2	YES	1	50.0	NO	0.125	YES
21	11	3	NO	1	33.3	NO	0.292	NO

Since delta is no longer decreasing and we are below the desired percentage, we present next larger target. Delta is reset to 1.00.

22	10	7	YES	5	71.4	YES	0.089	YES
23	10	8	NO	5	62.5	NO	0.000	YES
24	10	9	YES	6	66.7	YES	0.042	NO
25	11	...	...	...	...	...	...	...

So far we have described how stimuli are presented in the bracket method. We now must describe the rules by which we form an estimate of the threshold acuity. The simple rule adopted here was to merely average the two most frequent target gap sizes. It was assumed that the method would converge after some number of trials to the presentation of the two stimuli which bracketed the true threshold, and that these two gap sizes would be the most frequent. It then made sense to average these two most frequent gap sizes to produce an estimate of the true threshold. A more sophisticated method of forming an estimate from the bracket method utilizing the maximum likelihood approach will be described in a future report.

SIMULATION RESULTS FOR THE BRACKET METHOD

Figure 6 presents one simulation run of the bracket method. It displays the successful operation of the method as it quickly settled on the two gap sizes which bracketed the true threshold, and remained with these gap sizes for the duration of the run. The estimate of 0.5012 mva was calculated by averaging the total number of times gap sizes #10 and #11 were presented. In this example there were no stimuli excluded from the calculation since the two most frequent stimuli included all the stimulus presentations.

Figure 7 is a graph which repeats the curves from Figure 4 and includes the theshold acuity estimates from one hundred simulation runs of the bracket method. The bracket method is easily seen to provide a closer estimate to the true threshold acuity than either of the two estimators from the up and down method. This is certainly one feature in its favor. But how does the bracket method fare when compared with the up and down method when the variability of the estimates is the issue? Figure 8 answers this question.

Figure 8 compares the standard deviation of the one hundred estimates of the threshold acuity for the bracket method and the up and down method. The error bars extending from each mean value represent plus or minus two standard deviations. Data are shown at 10, 50, and 90 trials for the bracket method and at 4, 14, and 20 crossings for the up and down method. In each case the range of estimates from the bracket method is considerably smaller than the corresponding range for the up and down method. This illustrates the fact that there is less inherent variability in the bracket method as opposed to the up and down method, and therefore serves as a more precise measuring instrument for determining visual acuity thresholds.

To what extent would our conclusions about the two methods be affected if the simulation model were changed? Although not all the pertinent changes to the simulation model were investigated in this research effort, the results of allowing the standard deviation of the psychometric curve to vary are shown in Figure 9. The visual effect on the psychometric curve of allowing the standard deviation of the curve to get smaller is a steepening of the curve, while, conversely, allowing the standard deviation to increase results in a flattening of the curve.

Up to this point, all the simulation runs have been conducted with a value of 0.08 mva for the standard deviation of the psychometric curve. To observe the effect upon the average threshold estimate and the variability of the threshold estimate, the standard deviation of the psychometric curve was studied for values of 0.01 mva to 0.21 mva in steps of 0.01 mva. The results are shown in Figure 9. Three representative points are shown for both the bracket and the up and down method. Each point and error bars result from one hundred simulation runs at the standard deviation specified on the x-axis. The range of estimates for the bracket method always remains smaller than the up and down method for all values of the standard deviation of the psychometric curve. As we have observed previously, the average of the bracket method estimates is closer to the true threshold than the up and down method over all values of the standard deviation of the psychometric curve.

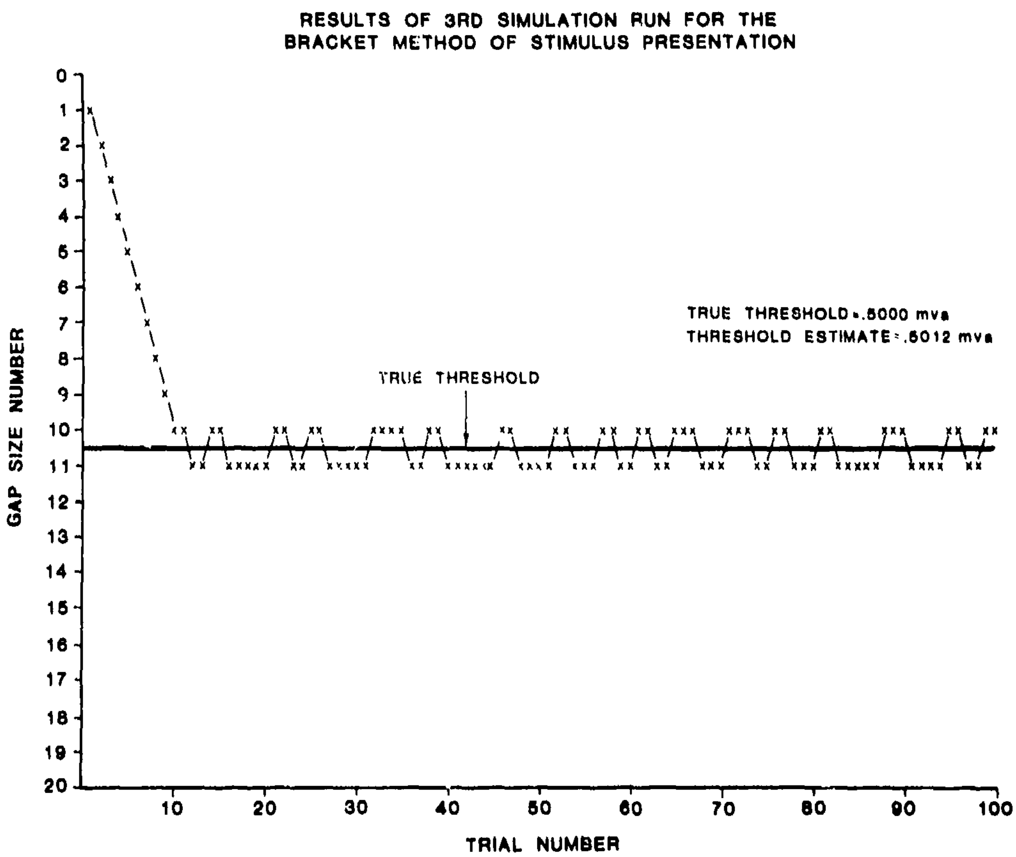

Figure 6

An example of the stimuli presented in a 100 trial simulation run of the bracket method. The stimuli presented do "bracket" the true threshold.

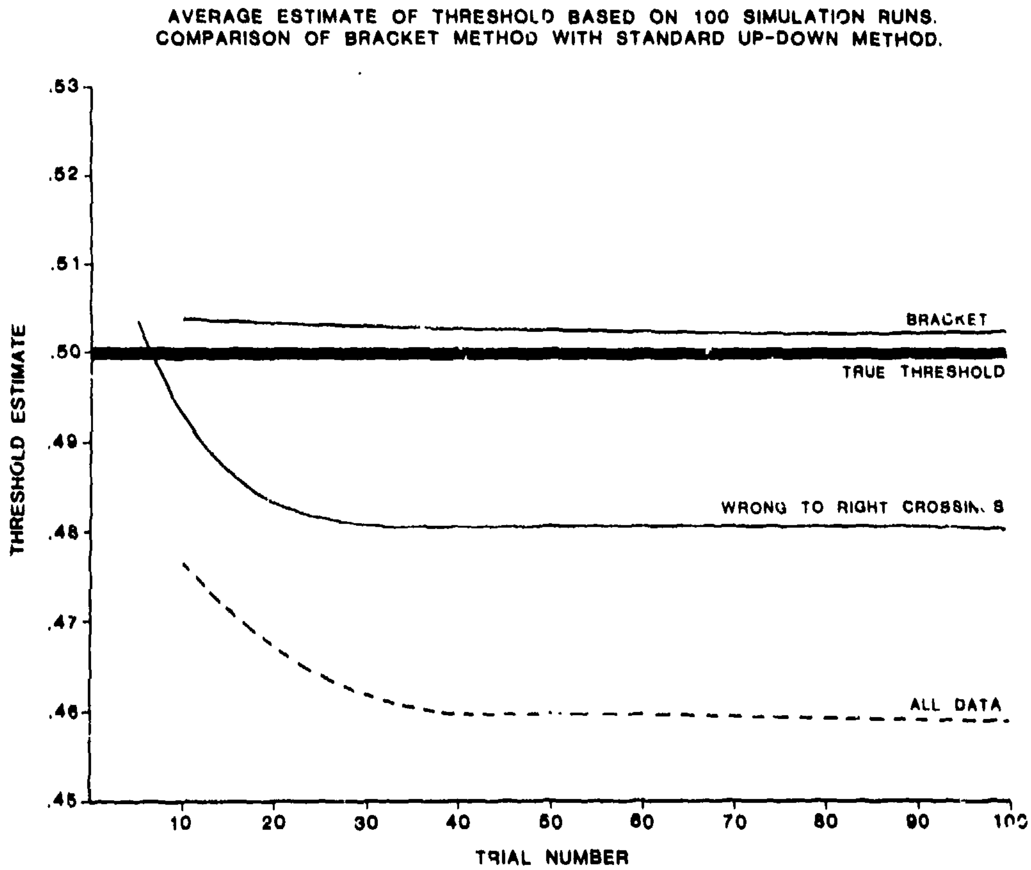

Figure 7

The bracket method estimator is closer to the true threshold at any trial number than is either of the two estimators from the up-down method.

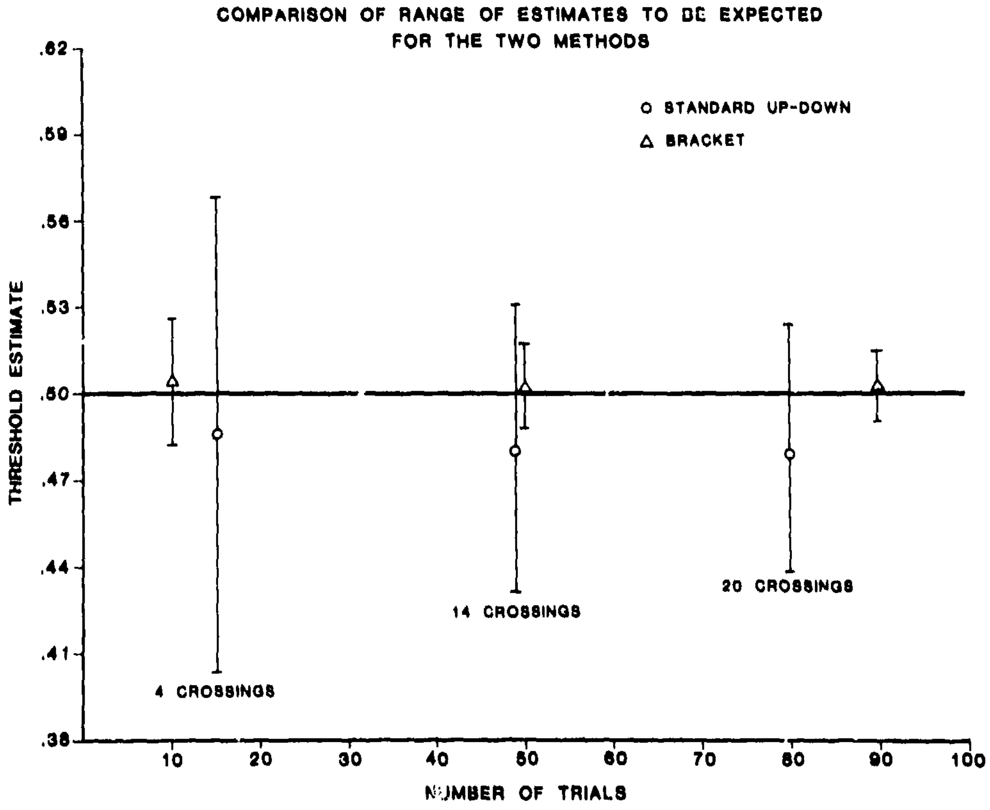

Figure 8

The variability of the threshold estimate when the bracket method is employed is smaller than the variability of the threshold estimate from the up-down method. This relationship holds over any number of trials.

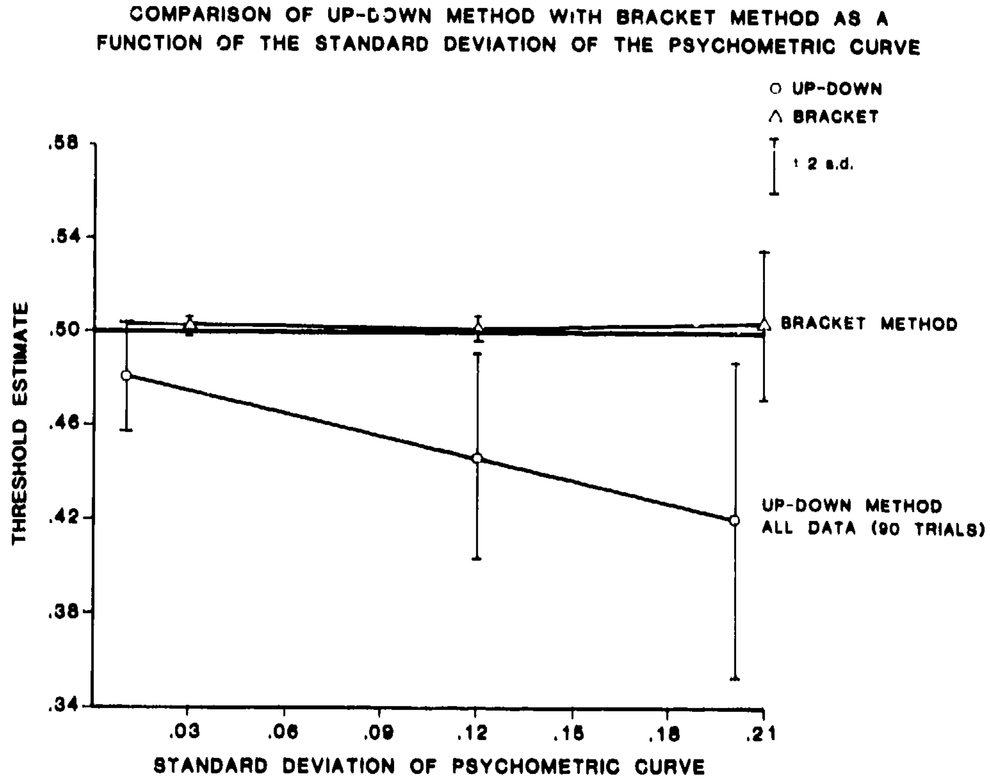

Figure 9

The behavior of the threshold estimates from the bracket method and the "all data estimator" from the up-down method when one of the parameters (σ) of the psychometric curve is allowed to vary.

VIOLATIONS OF THE MODEL

I will indicate here two points where the model presented is likely to be violated when testing actual subjects.

1) It is likely that there will be some session to session variability in the threshold acuity for actual subjects as a function of such factors as practice and fatigue.

2) There is some empirical evidence that threshold acuity can shift even during the course of a run.

The model presented here assumed that the parameters of the psychometric curve remained constant during the course of a simulation run as well as from run to run. The net result of both of the above situations in the testing of actual subjects is that the standard deviation of the estimates will be inflated compared to the results shown in this paper. Ofcourse, since the bracket method has the lower inherent variability, it would be the more sensitive instrument for detecting the occurrence of shifts in the threshold acuity due to the situations described above.

CONCLUSION

An accurate and economical method for determining visual acuity threshold was necessary for many tests in the VTB. The classical up and down method was considered as the method of choice to accomplish this task. However, there were certain unanswered questions as to how this method would perform with a four alternative forced choice task, and how large the resulting variability of the estimator would be.

To resolve these questions a mathematical model of how subjects might emit responses in the up and down method was constructed. This model was run on the computer with parameters chosen to characterize one of the acuity tests in the VTB. The intent of the computer simulation was to generate a relatively large sample of estimates of the threshold acuity for the up and down method.

The two statistics of interest from these computer generated samples were: 1) the average estimate of the threshold acuity and 2) the standard deviation of this sample of estimates. The first statistic was to judge how much the four alternative forced choice nature of the task had biased the estimate of a known threshold acuity. The second statistic gave some idea of the size of the error one should attach to an estimate of the threshold acuity when using either of the sequential presentation strategies discussed.

During the course of this research, a new method of presenting stimuli was also examined with regard to these two statistics. The estimates from this new method, called the bracket method, proved to be superior for both statistical criteria, even when an important parameter of the underlying model generating the responses was varied. Two possible violations of the model used in the simulation which might occur during the testing of actual subjects were discussed.

REFERENCES

1. Blower, D. J., The bias in the presentation of stimuli when the up and down method is used with forced choice responding. NAMRL-1269. Pensacola, Fl: Naval Aerospace Medical Research Laboratory, July, 1980.

2. Brownlee, K. A., Hodges, J. L., and Rosenblatt, M., The up and down method with small samples. J. Am. Stat. Assoc., 48:262-277, 1953.

3. Cornsweet, T. N., The staircase method in psychophysics. Am. J. of Psych., 75:485-491, 1962.

4. Dixon, W. J., and Mood, A. M., A method for obtaining and analyzing sensitivity data. J. Am. Stat. Assoc., 43:109-126, 1948.

5. Levitt, H., Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am., 49:467-477, 1970.

6. Lythgoe, R. J., The measurement of visual acuity. Med. Res. Coun. Spec. Rep., Ser 173, 1-85, London, England, 1932.

7. Pentland, A., Maximum likelihood estimation: The best PEST. Perception & Psychophysics, 28:377-379, 1980.

8. Taylor, M. M., and Creelman, C. D., PEST: Efficient estimates on probability functions. J. Acoust. Soc. Am., 41:782-787, 1967.

9. Wetherill, G. B., Chen, H., and Vasudeva, R. B., Sequential estimation of quantal response curves: A new method of estimation. Biometrika, 53:439-454, 1966.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NAMRL- 1282	2. GOVT ACCESSION NO. AD-A111 821	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DETERMINING VISUAL ACUITY THRESHOLDS: A Simulation Study of Stimulus Presentation Strategies		5. TYPE OF REPORT & PERIOD COVERED Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David J. Blower, LT MSC USN		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Aerospace Medical Research Laboratory Naval Air Station, Pensacola, Florida 32508		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NMRDC ZF51.524-004-2011 NAVAIRSYSCOM W 43-13.8881
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research and Development Command Bethesda, Maryland 20014		12. REPORT DATE 10 August 1981 13. NUMBER OF PAGES 21
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Visual acuity thresholds
Computer simulation
Statistical estimates
Stimulus presentation
Up and down method
Bracket method

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An accurate and economical method for determining visual acuity thresholds was necessary for many tests in the Vision Test Battery (VTB). There were certain unanswered questions as to how the classical up and down method of presenting stimuli would perform in this regard with a four alternative forced choice task. To answer these questions computer simulation runs of the up and down method were carried out. A model of the psychometric function relating probability of correct detection to stimulus gap size was based on the cumulative normal distribution. The parameters of this model were chosen to characterize one of the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) acuity tests in the VTB. The intent of the simulation with this model was to generate a large sample of acuity thresholds, and, from this sample, form an estimate of the average acuity threshold and its variability for any given number of trials.

An alternative method of presenting stimuli in a sequential manner was also studied via simulation. This method, called the bracket method, proved to be superior to the up and down method in calculating an average estimate of threshold acuity. The error attached to the threshold acuity estimate was also smaller in the bracket method. When the slope of the psychometric curve was varied over a large range, the bracket method retained these advantages over the up and down method.